TECHNICAL WORKING PAPER SERIES

TWO-SAMPLE INSTRUMENTAL VARIABLES ESTIMATORS

Atsushi Inoue
Gary Solon

Technical Working Paper 311
http://www.nber.org/papers/T0311

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2005

Two-Sample Instrumental Variables Estimators
Atsushi Inoue and Gary Solon
NBER Technical Working Paper No. 311
June 2005
JEL No. C3

ABSTRACT

Following an influential article by Angrist and Krueger (1992) on two-sample instrumental variables (TSIV) estimation, numerous empirical researchers have applied a computationally convenient two-sample two-stage least squares (TS2SLS) variant of Angrist and Krueger's estimator. In the two-sample context, unlike the single-sample situation, the IV and 2SLS estimators are numerically distinct. Our comparison of the properties of the two estimators demonstrates that the commonly used TS2SLS estimator is more asymptotically efficient than the TSIV estimator and also is more robust to a practically relevant type of sample stratification.

Atsushi Inoue
Department of Agricultural and Resource Economics
North Carolina State University
Raleigh, NC 27607-8109
atsushi.inoue@ncsu.edu

Gary Solon
Department of Economics
University of Michigan
Ann Arbor, MI 48109-1220
and NBER
gsolon@umich.edu

Two-Sample Instrumental Variables Estimators

I. Introduction

A familiar problem in econometric research is consistent estimation of the coefficient vector in the linear regression model

$$y = W\theta + \varepsilon \tag{1}$$

where y and ε are $n \times 1$ vectors and W is an $n \times k$ matrix of regressors, some of which are endogenous, i.e., contemporaneously correlated with the error term ε. As is well known, the ordinary least squares estimator of θ is inconsistent, but consistent estimation is still possible if there exists an $n \times q$ ($q \geq k$) matrix Z of valid instrumental variables. For example, in the case of exact identification with $q = k$, the conventional instrumental variables (IV) estimator is

$$\hat{\theta}_{IV} = (Z'W)^{-1}Z'y. \tag{2}$$

With exact identification, this estimator is identical to the two-stage least squares (2SLS) estimator

$$\hat{\theta}_{2SLS} = (\hat{W}'\hat{W})^{-1}\hat{W}'y \tag{3}$$

where $\hat{W} = Z(Z'Z)^{-1}Z'W$. If, in addition, ε is i.i.d. normal, this estimator is asymptotically efficient among "limited information" estimators.

An influential article by Angrist and Krueger (1992) has pointed out that, under certain conditions, consistent instrumental variables estimation is still possible even when only y and Z (but not W) are observed in one sample and only W and Z (but not y) are observed in a second distinct sample. In that case, the same moment conditions that lead to the conventional IV estimator in equation (2) motivate the "two-sample instrumental variables" (TSIV) estimator

$$\hat{\theta}_{TSIV} = (Z_2'W_2/n_2)^{-1}(Z_1'y_1/n_1) \tag{4}$$

where Z_1 and y_1 contain the n_1 observations from the first sample and Z_2 and W_2 contain the n_2 observations from the second.

Of the many empirical researchers who have since used a two-sample approach (e.g., Bjorklund and Jantti, 1997; Currie and Yelowitz, 2000; Dee and Evans, 2003; Borjas, 2004), nearly all have used the "two-sample two-stage least squares" (TS2SLS) estimator

$$\hat{\theta}_{TS2SLS} = (\hat{W}_1'\hat{W}_1)^{-1}\hat{W}_1'y_1 \tag{5}$$

where $\hat{W}_1 = Z_1(Z_2'Z_2)^{-1}Z_2'W_2$. These researchers may not have been aware that the equivalence of IV and 2SLS estimation in a single sample does not carry over to the two-sample case. Instead, it is easy to show that, in the exactly identified case,

$$\hat{\theta}_{TS2SLS} = (Z_2'W_2/n_2)^{-1}C(Z_1'y_1/n_1) \tag{6}$$

where $C = (Z_2'Z_2/n_2)(Z_1'Z_1/n_1)^{-1}$ can be viewed as a sort of correction factor for differences between the two samples in their covariance matrices for Z. Under Angrist and Krueger's assumptions, those differences would disappear asymptotically. As a result, the correction matrix C would converge in probability to the identity matrix, and the TSIV and TS2SLS estimators would have the same probability limit. In finite samples, however, the TSIV estimator originally proposed by Angrist and Krueger and the TS2SLS estimator typically used by practitioners are numerically distinct estimators.[1]

[1] In a subsequent paper on split-sample IV estimation as a method for avoiding finite-sample bias when the instruments are only weakly correlated with the endogenous regressors, Angrist and Krueger (1995) noted the distinction between TS2SLS and TSIV and conjectured (incorrectly) that the two estimators have the same asymptotic distribution. In another related literature, on "generated regressors," first-stage estimation is performed to create a proxy for an unobserved regressor in the second-stage equation, rather than to treat the endogeneity of the regressor. Murphy and Topel (1985) discussed the instance in which the first-stage estimation is based on a different sample than the second-stage estimation.

The obvious question then becomes: Which estimator should be preferred? Our paper addresses this question while going beyond the simple example described above to consider overidentified as well as exactly identified models, heteroskedastic errors, and stratified samples. It turns out that the two-sample two-stage least squares approach commonly used by practitioners not only is computationally convenient, but also has theoretical advantages. Its implicit correction for differences between the two samples in the distribution of Z yields a gain in asymptotic efficiency and also maintains consistency in the presence of a practically relevant form of stratified sampling.

The outline of this paper is as follows. Section II compares the asymptotic efficiency of the TSIV and TS2SLS estimators in a basic model. Section III considers departures from the basic model, such as heteroskedasticity and stratification. Section IV summarizes and concludes the paper.

II. Asymptotic Efficiency

In this section, we will compare two-sample IV estimators in a general single-equation framework:

$$y_{1i} = \beta' x_{1i} + \gamma' z_{1i}^{(1)} + \varepsilon_{1i} = \theta' w_{1i} + \varepsilon_{1i}, \tag{7}$$

$$x_{1i} = \Pi z_{1i} + \eta_{1i}, \tag{8}$$

$$x_{2i} = \Pi z_{2i} + \eta_{2i}, \tag{9}$$

where x_{1i} and x_{2i} are p-dimensional random vectors, $z_{1i} = [z_{1i}^{(1)'} \ z_{1i}^{(2)'}]'$ and z_{2i} are $q(= q^{(1)} + q^{(2)})$-dimensional random vectors, w_{1i} is a $k(= p + q^{(1)})$-dimensional random vector, and Π is a $p \times q$ matrix of parameters.

For efficiency comparison, it is useful to characterize these estimators as generalized method of moments (GMM) estimators. First the TSIV estimator is a GMM estimator based on moment conditions

$$E\left[z_{1i}(y_{1i} - z_{1i}^{(1)'}\gamma) - z_{2i}x_{2i}'\beta\right] = 0. \tag{10}$$

Next the TS2SLS estimator is a GMM estimator based on

$$
\begin{aligned}
E[z_{1i}(y_{1i} - z_{1i}'\Pi'\beta - z_{1i}^{(1)'}\gamma)] &= 0, & (11) \\
E[z_{2i} \otimes (x_{2i} - \Pi z_{2i})] &= 0. & (12)
\end{aligned}
$$

When Π is defined to be the coefficient on z_i in the population linear projection of x_i on z_i in the second sample, (12) always holds by definition of linear projections.

Finally we consider the two-sample limited-information maximum likelihood (TSLIML) estimator for efficiency comparison. Let $\sigma_{11} = E[(\varepsilon_{1i} + \beta'\eta_{1i})^2]$ and $\Sigma_{22} = E(\eta_{2i}\eta_{2i}')$. When $[\varepsilon_i \ \eta_{1i}']'$ and η_{2i} are normally distributed the log of the likelihood function can be written as

$$
\begin{aligned}
\ln L = & -\frac{n}{2}\ln(2\pi) - \frac{n_1}{2}\ln(\sigma_{11}) - \frac{n_2}{2}\ln|\Sigma_{22}| \\
& -\frac{1}{2\sigma_{11}}\sum_{i=1}^{n_1}(y_{1i} - \beta'\Pi z_{1i} - \gamma' z_i^{(1)})^2 \\
& -\frac{1}{2}\sum_{i=1}^{n_2}(x_{2i} - \Pi z_{2i})'\Sigma_{22}^{-1}(x_{2i} - \Pi z_{2i}).
\end{aligned}
$$

The TSLIML estimator is asymptotically equivalent to a GMM estimator based on the population first-order conditions for the TSLIML estimator:

$$
\begin{aligned}
E[\Pi z_{1i}(y_{1i} - \beta'\Pi z_{1i} - \gamma' z_i^{(1)})] &= 0, & (13) \\
E[z_i^{(1)}(y_{1i} - \beta'\Pi z_{1i} - \gamma' z_i^{(1)})] &= 0, & (14) \\
E(z_{1i} \otimes \beta u_{1i}/\sigma_{11} + z_{2i} \otimes \Sigma_{22}^{-1}\eta_{2i}) &= 0, & (15) \\
E(u_{1i}^2/\sigma_{11}^2 - 1/\sigma_{11}) &= 0, & (16) \\
E[(\Sigma_{22}^{-1}\eta_{2i}) \otimes (\Sigma_{22}^{-1}\eta_{2i}) - |\Sigma_{22}|\mathrm{tr}(\Sigma_{22}^{-1})]D_2 &= 0, & (17)
\end{aligned}
$$

where D_2 is a $p^2 \times p(p+1)/2$ matrix such that $\text{vec}(\Sigma_{22}) = D_2 \text{vech}(\Sigma_{22})$.

To derive the asymptotic distributions of these estimators we assume the following conditions.

Assumption 1.

(a) $\{[y_{1i}, z_{1i}']'\}_{i=1}^{n_1}$ and $\{[x_{2i}, z_{2i}']'\}_{i=1}^{n_2}$ are iid random vectors with finite fourth moments and are independent.

(b) $E(\varepsilon_{1i}|z_{1i}) = 0$, $E(\eta_{1i}|z_{1i}) = 0$ and $E(\eta_{2i}|z_{2i}) = 0$.

(c) $E(u_{1i}^2|z_{1i}) = \sigma_{11}$ and $E(\eta_{2i}\eta_{2i}'|z_{2i}) = \Sigma_{22}$ where $u_{1i} = \varepsilon_{1i} + \beta'\eta_{1i}$, $\sigma_{11} > 0$ and Σ_{22} is positive definite.

(d) Third moments of $[\varepsilon_{1i}\ \eta_{1i}']$ and those of η_{2i} are all zero conditional on z_{1i} and z_{2i}, respectively.

(e) For the TSIV estimator

$$\text{rank}\begin{bmatrix} E(z_{2i}x_{2i}') & 0 \\ 0 & E(z_{1i}z_{1i}^{(1)\prime}) \end{bmatrix} = \dim(\theta)$$

and for the TS2SLS and TSLIML estimators $\text{rank}[E(z_{1i}w_{1i}')] = \dim(\theta)$.

(f) $E(z_{1i}z_{1i}')$ and $E(z_{2i}z_{2i}')$ are nonsingular.

(g) $E(z_{1i}x_{1i}') = E(z_{2i}x_{2i}') = E(z_i x_i')$ and $E(z_{1i}z_{1i}') = E(z_{2i}z_{2i}') = E(z_i z_i')$.

(h) $\lim_{n_1,n_2\to\infty} n_1/n_2 = \kappa$ for some $\kappa > 0$.

Remarks. Assumption (c) rules out conditional heteroskedasticity, which will be considered in Section III. Assumption (d) is used to simplify the derivation of the asymptotic covariance matrix of the TSLIML estimator. Assumption (g) provides a basis for combining two samples.[2]

Proposition 1. Under Assumption 1, $\hat{\theta}_{TSIV}$, $\hat{\theta}_{TS2SLS}$ and $\hat{\theta}_{TSLIML}$ are $\sqrt{n_1}$-consistent[3] and asymptotically normally distributed with asymptotic covariance matrices Σ_{TSIV}, Σ_{TS2SLS} and Σ_{TSLIML}, respectively, where

$$\Sigma_{TSIV} = \left\{E(w_i z_i') \left[(\sigma_{11} + \kappa\beta'\Sigma_{22}\beta) E(z_i z_i') + Cov(z_{1i}z_{1i}'\Pi\beta) + \kappa Cov(z_{2i}z_{2i}'\Pi\beta)\right]^{-1} E(z_i w_i')\right\}^{-1} \quad (18)$$

$$\Sigma_{TS2SLS} = \{E(w_i z_i')[(\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)E(z_i z_i')]^{-1}E(z_i w_i')\}^{-1}, \quad (19)$$

$$\Sigma_{TSLIML} = \Sigma_{TS2SLS}, \quad (20)$$

and $Cov(z_i z_i'\Pi\beta) = E(z_i z_i'\Pi\beta\beta'\Pi' z_i z_i') - E(z_i z_i'\Pi\beta)E(\beta'\Pi' z_i z_i')$.

Remarks. 1. In the notation of Angrist and Krueger (1992),

$$\phi_1 = \sigma_{11}E(z_i z_i') + Cov(z_i z_i'\Pi\beta),$$

$$\omega_2 = \beta'\Sigma_{22}\beta E(z_i z_i') + Cov(z_{2i}z_{2i}'\Pi\beta).$$

2. Since $Cov(z_{1i}z_{1i}'\Pi\beta) + \kappa Cov(z_{2i}z_{2i}'\Pi\beta)$ is positive semidefinite, it follows that $\Sigma_{TSIV} - \Sigma_{TS2SLS}$ is positive semidefinite. Thus, the TS2SLS estimator is more efficient than the TSIV estimator. The asymptotic efficiency gain comes from the implicit correction of the TS2SLS estimator for differences between the finite-sample distributions of z_{1i} and z_{2i}. 3. Proposition 1 shows that the TS2SLS and TSLIML estimators are asymptotically equivalent. It follows that, when the disturbance terms $[\varepsilon_i\ \eta_i']'$ are jointly normally distributed, the TS2SLS estimator is asymptotically efficient within the

[2]One can show that the TS2SLS estimator requires a weaker condition $E(z_{1i}x_{1i}') = cE(z_{2i}x_{2i}')$ and $E(z_{1i}z_{1i}') = cE(z_{2i}z_{2i}')$ for some c. Because c does not have to be unity, the TS2SLS estimator is more robust than the TSIV estimator.

[3]Following Angrist and Krueger (1992), we scale the estimator by $\sqrt{n}_1$.

class of “limited information” estimators.[4]

As an illustration, let us consider the case in which one endogenous variable is the only explanatory variable and we have one instrument, i.e., $p = q = q^{(2)} = 1$. In this case one can show that

$$\sqrt{n_1}(\hat{\beta}_{TSIV} - \beta) - \sqrt{n_1}(\hat{\beta}_{TS2SLS} - \beta) = \frac{n_1^{-1/2} \sum_{i=1}^{n_1} z_{1i}^2 - \sqrt{\kappa} n_1^{-1/2} \sum_{i=1}^{n_2} z_{2i}^2}{(1/n_2) \sum_{i=1}^{n_2} z_{2i} x_{2i}} \Pi\beta + o_p(1).$$

The first term on the RHS will be asymptotically independent of $\sqrt{n_1}(\hat{\beta}_{TS2SLS} - \beta)$ and have a positive variance even asymptotically and it follows from Proposition 1 that its variance is given by

$$\frac{Var(z_{1i}^2 \Pi\beta) + \kappa Var(z_{2i}^2 \Pi\beta)}{[E(z_i x_i)]^2}.$$

III. Robustness to Departures from the Basic Model

In this section, we consider departures from the basic model, namely, conditional heteroskedasticity and stratification. In doing so, we assume that the inverse of a consistent estimator of the variance covariance matrix of moment conditions is used as an optimal weighting matrix to achieve efficiency among GMM estimators given the moment conditions. We will call the resulting GMM estimator based on (10) the TSIV estimator and denote it by $\tilde{\theta}_{TSIV}$, and the resulting GMM estimator based on (11) and (12) will be called the TS2SLS estimator and be denoted by $\tilde{\theta}_{TS2SLS}$.

[4]In Monte Carlo experiments, we have verified that these asymptotic results accurately characterize the finite-sample behavior of the TSIV, TS2SLS, and TSLIML estimators. The exception is that, when the instruments are very weakly correlated with the endogenous regressor, all three estimators appear to be biased towards zero. This corroborates an analytical result of Angrist and Krueger (1995) concerning TS2SLS. The finding that TSLIML is subject to a similar finite-sample bias is interesting. Apparently, the well-known tendency for LIML to be less biased than 2SLS when the instruments are weak in the single-sample setting (e.g., Angrist, Imbens, and Krueger, 1999) does not carry over to the two-sample setting.

First, consider the case of conditional heteroskedasticity. We replace Assumption (c) by

(c') $E(u_{1i}^2 z_{1i} z_{1i}') > 0$ and $E(z_{2i} z_{2i}' \otimes \eta_{2i} \eta_{2i}')$ is positive definite.

Then we obtain the following:

Proposition 2. Under Assumption 1 with Assumption (c) replaced by Assumption (c'), $\tilde{\theta}_{TSIV}$ and $\tilde{\theta}_{TS2SLS}$ are consistent and asymptotically normally distributed with asymptotic covariance matrices Σ_{TSIV}^{hetero} and Σ_{TS2SLS}^{hetero}, respectively, where

$$\Sigma_{TSIV}^{hetero} = \left\{E(w_i z_i') \left[E(u_i^2 z_i z_i') + \kappa E(\beta' \eta_i \eta_i' \beta z_i z_i') + Cov(z_{1i} z_{1i}' \Pi \beta) + \kappa Cov(z_{2i} z_{2i}' \Pi \beta)\right]^{-1} E(z_i w_i')\right\}^{-1} \tag{21}$$

$$\Sigma_{TS2SLS}^{hetero} = \left\{E(w_i z_i') \left[E(u_i^2 z_i z_i') + \kappa E(\beta' \eta_i \eta_i' \beta z_i z_i')\right]^{-1} E(z_i w_i')\right\}^{-1}. \tag{22}$$

Remark. As in Proposition 1, the GMM estimator based on (11) and (12) is asymptotically more efficient than the GMM estimator based on (10).

Next consider a practically relevant type of stratified sampling. Suppose that either or both of the two samples use sampling rates that vary with some of the instrumental variables. For example, household surveys commonly use different sampling rates by race or location, which may be among the regressors in $z_{1i}^{(1)}$ in equation (7). The National Longitudinal Surveys have oversampled African-Americans, the Health and Retirement Study has oversampled residents of Florida, and the Current Population Survey has oversampled in less populous states.

When analyzing stratification, it is useful to define two binary selection variables:

$$s_{1i} = \begin{cases} 1 & \text{if the first sample includes the } i\text{th observation} \\ 0 & \text{otherwise} \end{cases}$$

$$s_{2i} = \begin{cases} 1 & \text{if the second sample includes the } i\text{th observation} \\ 0 & \text{otherwise} \end{cases}$$

By construction, $s_{1i} + s_{2i} = 1$ for each i. Note that the TSIV estimator and the TS2SLS estimator can be viewed as GMM estimators based on population moment conditions

$$E\left[\frac{s_{i1}}{E(s_{1i})} z_i(y_i - z_{1i}'\gamma) - \frac{s_{2i}}{E(s_{2i})} z_i x_i'\beta\right] \;=\; 0, \tag{23}$$

and

$$E[s_{1i} z_i(y_i - z_{1i}'\gamma) - \kappa s_{2i} z_i x_i'\beta] \;=\; 0, \tag{24}$$

$$E[(s_{1i} - \kappa s_{2i})\mathrm{vech}(z_i z_i')] \;=\; 0, \tag{25}$$

respectively.

Assumption 2.

(a) $[s_{1i}, s_{2i}, x_i, y_i, z_i']'$ is an iid random vector with finite fourth moments.

(b) $E(\varepsilon_i|s_{1i}, z_i) \;=\; 0$ and $E(\eta_i|s_{1i}, s_{2i}, z_i) \;=\; 0$.

(c) $E(u_i^2|s_{1i}, z_i) = \sigma_{11}$ and $E(\eta_i\eta_i'|s_{2i}, z_i) = \Sigma_{22}$ where $u_i = \varepsilon_i + \beta'\eta_i$, $\sigma_{11} > 0$ and Σ_{22} is positive definite.

(d) For the TSIV estimator

$$\mathrm{rank}\begin{bmatrix} E(s_{2i} z_i x_i') & 0 \\ 0 & E(s_{1i} z_i z_i^{(1)'}) \end{bmatrix} \;=\; \dim(\theta)$$

and for the TS2SLS and TSLIML estimators $\mathrm{rank}[E(s_{1i} z_i w_i')] \;=\; \dim(\theta)$.

(e) $E(s_{1i} z_i z_i')$ and $E(s_{2i} z_i z_i')$ are nonsingular.

(f) $E(s_{1i} s_{2i}) \;=\; 0$.

(g) s_{1i} and s_{2i} are independent of x_i.

(h) $E((s_{1i}/p_1)z_i x_i') = \kappa E((s_{2i}/p_2)z_i x_i')$ and $E((s_{1i}/p_1)z_i z_i') = \kappa E((s_{2i}/p_2)z_i z_i')$ for some constant κ where $p_1 = P(s_{1i} = 1)$ and $p_2 = P(s_{2i} = 1)$.

Remarks.

Because

$$
\begin{array}{lclcl}
E[s_{1i}z_i(y_i - z_i'\Pi'\beta - z_{1i}'\gamma)] & = & E(s_{1i}z_i\varepsilon_i) + E[s_{1i}z_i\eta_i']\beta & = & 0, \\
E[s_{2i}z_i \otimes (x_i - \Pi z_i)] & = & E(s_{2i}z_i \otimes \eta_i) & = & 0,
\end{array}
$$

this type of stratification does not affect the validity of the moment conditions for the TS2SLS estimator. In contrast, when the two samples differ in their stratification schemes, the population moment function for the TSIV estimator

$$
\begin{aligned}
E\left[\frac{s_{1i}}{p_1}z_i y_i - \frac{s_{2i}}{p_2}z_i x_i'\beta\right] &= \frac{1}{p_1}E(s_{1i}z_i\varepsilon_i) + E\left[\left(\frac{s_{1i}}{p_1} - \frac{s_{2i}}{p_2}\right) z_i x_i'\right]\beta \\
&= E\left[\left(\frac{s_{1i}}{p_1} - \frac{s_{2i}}{p_2}\right) z_i z_i'\right]\Pi'\beta \qquad (26)
\end{aligned}
$$

is likely to be nonzero. As a result, the TSIV estimator will not be consistent in general.

It is possible, however, to modify the TSIV estimator so that it is robust to stratification. Define a robust modification of the TSIV estimator, $\tilde{\theta}_{RTSIV} = [\tilde{\beta}_{RTSIV}' \ \tilde{\gamma}_{RTSIV}']'$, by a GMM estimator based on sample moment functions

$$
E[s_{1i}z_i(y_i - z_{1i}'\gamma) - \kappa s_{2i}z_i x_i'\beta] = 0, \qquad (27)
$$

$$
E[(s_{1i} - \kappa s_{2i})\mathrm{vech}(z_i z_i')] = 0. \qquad (28)
$$

Proposition 3. Under Assumption 2, $\tilde{\theta}_{TS2SLS}$ and $\tilde{\theta}_{RTSIV}$ are consistent and asymptotically normally distributed with asymptotic covariance matrices Σ_{TS2SLS} and Σ_{RTSIV}, respectively, where

$$
\Sigma_{TS2SLS} = (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)[E(s_{1i}w_i z_i')E(s_{1i}z_i z_i')^{-1}E(s_{1i}z_i w_i')]^{-1} \qquad (29)
$$

and Σ_{RTSIV} is the upper-left $k \times k$ submatrix of $(G'_{RTSIV} V_{RTSIV}^{-1} G_{RTSIV})^{-1}$,

$$G_{RTSIV} = -\begin{bmatrix} E(s_{1i} z_i w_i') & E(s_{2i} z_i x_i')\beta \\ 0 & E(s_{2i}\text{vech}(z_i z_i')) \end{bmatrix},$$

$$V_{RTSIV} = \begin{bmatrix} (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)E(s_{1i}z_iz_i') + E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\beta'\Pi z_iz_i'] & E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\text{vech}(z_iz_i')'] \\ E[(s_{1i} + \kappa^2 s_{2i})\text{vech}(z_iz_i')\beta'\Pi z_iz_i'] & E[(s_{1i} + \kappa^2 s_{2i})\text{vech}(z_iz_i')\text{vech}(z_iz_i')'] \end{bmatrix}.$$

Moreover, $\Sigma_{TS2SLS} \leq \Sigma_{RTSIV}$.[5]

Remarks. The TS2SLS estimator is more efficient than the robust version of the TSIV estimator. Proposition 3 does not rule out cases in which $\Sigma_{TS2SLS}^{strat} = \Sigma_{RTSIV}^{strat}$. Even in those cases, the TS2SLS estimator may be still preferable for the following reason. Note that the number of overidentifying restrictions for the robust TSIV estimator is $q_2 + q(q+1)/2 - p - 1$ whereas the one for the TS2SLS estimator is $q_2 - p$. In typical applications, $q > 1$ and thus $q_2 + q(q+1)/2 - p - 1$ is greater than $q_2 - p$. Because using too many moment conditions often results in poor finite-sample performance of the GMM estimator (e.g., Tauchen, 1986, and Andersen and Sorensen, 1996), the TS2SLS estimator may be preferable in small samples.

IV. Summary

Following Angrist and Krueger's (1992) influential work on two-sample instrumental variables (TSIV) estimation, many applied researchers have used a two-sample two-stage least squares (TS2SLS) variant of Angrist and Krueger's estimator. In the two-sample context, unlike the single-sample setting, the IV and 2SLS estimators are numerically distinct. Under the conditions in which both estimators are consistent, we have shown that the commonly used TS2SLS approach is more asymptotically efficient because it implicitly corrects for differences in the empirical distributions of the instrumental variables between the two samples. That correction also protects the TS2SLS

[5]Following the convention, $\Sigma_{TS2SLS}^{strat} \leq \Sigma_{RTSIV}^{strat}$ if and only if $\Sigma_{RTSIV}^{strat} - \Sigma_{TS2SLS}^{strat}$ is positive semidefinite.

estimator from an inconsistency that afflicts the TSIV estimator when the two samples differ in their stratification schemes.

Appendix: Proofs

In Propositions 1, 2 and 3, the consistency and asymptotic normality of the GMM estimators follow from the standard arguments. Thus, we will focus on the derivation of asymptotic variances in the following proofs.

Proof of Proposition 1. Let G_{TSIV} and V_{TSIV} denote the Jacobian and covariance matrix, respectively, of the moment condition (10). Under Assumptions 1(c)(g)(h), we have

$$\begin{aligned} G_{TSIV} &= -E(z_i w_i'), \\ V_{TSIV} &= (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)\, E(z_i z_i') + Cov(z_{1i}z_{1i}'\Pi\beta) + \kappa Cov(z_{2i}z_{2i}'\Pi\beta). \end{aligned}$$

from which (18) follows.

Let G_{TS2SLS} and V_{TS2SLS} denote the Jacobian and covariance matrix, respectively, of the moment conditions (11) and (12). Because the Jacobian and covariance matrices of the moment functions are given by

$$\begin{aligned} G_{TS2SLS} &= -\begin{bmatrix} E(z_i w_i') & E(z_i z_i') \otimes \beta' \\ 0 & E(z_i z_i') \otimes I_p \end{bmatrix}, \\ V_{TS2SLS} &= \begin{bmatrix} \sigma_{11} E(z_i z_i') & 0 \\ 0 & \kappa E(z_i z_i') \otimes \Sigma_{22} \end{bmatrix}, \end{aligned}$$

respectively, the asymptotic covariance matrix of the TS2SLS estimator is the $k \times k$ upper-left submatrix of the inverse of

$$G_{TS2SLS}' V_{TS2SLS}^{-1} G_{TS2SLS} = \begin{bmatrix} \frac{1}{\sigma_{11}} E(w_i z_i')(E(z_i z_i'))^{-1} E(z_i w_i') & E(w_i z_i') \otimes \frac{\beta'}{\sigma_{11}} \\ E(z_i w_i') \otimes \frac{\beta}{\sigma_{11}} & E(z_i z_i') \otimes \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right) \end{bmatrix}.$$

Because the $k \times k$ upper-left submatrix of $(G_{TS2SLS}' V_{TS2SLS}^{-1} G_{TS2SLS})^{-1}$ is the inverse of

$$\begin{aligned} &\frac{1}{\sigma_{11}} E(w_i z_i')(E(z_i z_i'))^{-1} E(z_i w_i') - E(w_i z_i') \otimes \frac{\beta'}{\sigma_{11}} \left[E(z_i z_i') \otimes \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right)\right]^{-1} E(z_i w_i') \otimes \frac{\beta}{\sigma_{11}} \\ &= \left[\frac{1}{\sigma_{11}} - \frac{\beta'}{\sigma_{11}} \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right)^{-1} \frac{\beta}{\sigma_{11}}\right]^{-1} E(w_i z_i')(E(z_i z_i'))^{-1} E(z_i w_i') \end{aligned}$$

by Theorem 13 in Amemiya (1985, p. 460) and

$$\left[\frac{1}{\sigma_{11}} - \frac{\beta'}{\sigma_{11}}\left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right)^{-1}\frac{\beta}{\sigma_{11}}\right]^{-1} = \sigma_{11} + \kappa\beta'\Sigma_{22}\beta$$

by Theorem 0.7.4 of Horn and Johnson (1985, p.19), (19) follows.

Under the assumptions, one can show that the asymptotic distribution of $\hat{\theta}_{TSLIML}$ and the one of the TSLIML estimator for $[\sigma_{11}\ \mathrm{vech}(\Sigma_{22})']'$ are independent. Thus, we can focus on the moment conditions (13), (14) and (15). Under the stated assumptions, the Jacobian G_{TSLIML} and covariance matrix V_{TSLIML} of these moment conditions are the same and are given by

$$\left[\begin{array}{ccc} \Pi E(z_i z_i')\Pi'/\sigma_{11} & \Pi E(z_i z_{1i}')/\sigma_{11} & \Pi E(z_i z_i')\otimes\beta'/\sigma_{11} \\ E(z_{1i}z_i')\Pi'/\sigma_{11} & E(z_{1i}z_{1i}')/\sigma_{11} & E(z_{1i}z_i')\otimes\beta'/\sigma_{11} \\ E(z_i z_i')\Pi'\otimes\beta/\sigma_{11} & E(z_i z_{1i}')\otimes\beta/\sigma_{11} & E(z_i z_i')\otimes\left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right) \end{array}\right].$$

Since

$$\begin{aligned} & \left[\begin{array}{cc} \Pi E(z_i z_i')\Pi'/\sigma_{11} & \Pi E(z_i z_{1i}')/\sigma_{11} \\ E(z_{1i}z_i')\Pi'/\sigma_{11} & E(z_{1i}z_{1i}')/\sigma_{11} \end{array}\right] \\ & -\left[\begin{array}{c} \Pi E(z_i z_i')\otimes\beta'/\sigma_{11} \\ E(z_{1i}z_i')\Pi'\otimes\beta'/\sigma_{11} \end{array}\right]\left[E(z_i z_i')\otimes\left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1}\right)\right]^{-1}\left[\begin{array}{c} \Pi E(z_i z_i')\otimes\beta'/\sigma_{11} \\ E(z_{1i}z_i')\Pi'\otimes\beta'/\sigma_{11} \end{array}\right]' \\ = & \left(1 - \frac{1}{\sigma_{11}}\beta'(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa}\Sigma_{22}^{-1})^{-1}\beta\right)\left[\begin{array}{cc} \Pi E(z_i z_i')\Pi'/\sigma_{11} & \Pi E(z_i z_{1i}')/\sigma_{11} \\ E(z_{1i}z_i')\Pi'/\sigma_{11} & E(z_{1i}z_{1i}')/\sigma_{11} \end{array}\right] \end{aligned}$$

and

$$\beta'(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{k}\Sigma_{22}^{-1})^{-1}\beta = \kappa\beta'\Sigma_{22}\beta - (\kappa\beta'\Sigma_{22}\beta)^2 / (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta),$$

one can show that the upper-left $k \times k$ submatrix of $(G_{TSLIML}'V_{TSLIML}^{-1}G_{TSLIML})^{-1}$ can be written as (20).

Proof of Proposition 2. Note that

$$A^{-1} - A^{-1}BA^{-1} + A^{-1}B(A+B)^{-1}BA^{-1} = B^{-1}A(A+B)^{-1}BA^{-1} = (A+B)^{-1}, \tag{30}$$

where $A = E(u_i^2 z_i z_i')$ and $B = E(\beta'\eta_i\eta_i'\beta z_i z_i')$. The remainders of the proofs of (21) and (22) are analogous to those of (18) and (19), respectively, and thus they are omitted.

Proof of Proposition 3. The proof of (29) is analogous to the one of (19) and is omitted.

By Theorem 13 of Amemiya (1985, p.460), the upper-left $q \times q$ submatrix matrix and the upper-right $q \times (q(1+1)/2)$ submatrix of V_{RTSIV}^{-1} can be written as the inverse of

$$\begin{aligned} & (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)E(s_{1i}z_iz_i') + E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\beta'\Pi z_iz_i'] \\ & -E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\text{vech}(z_iz_i')']\{E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\beta'\Pi z_iz_i']\}^{-1} \\ & \times E[(s_{1i} + \kappa^2 s_{2i})\text{vech}(z_iz_i')\beta'\Pi z_iz_i'] \\ = \; & (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)E(s_{1i}z_iz_i') \end{aligned} \tag{31}$$

where the equality follows because the residuals from regressing $\beta'\Pi z_iz_i'$ on $\text{vech}(z_iz_i)'$ are numerically zero by the projection argument. The other submatrices of V_{RAK}^{-1} can be obtained by using the same theorem. After some matrix algebra, one can show that

$$(G_{RTSIV}'V_{RTSIV}^{-1}G_{RTSIV})^{-1} = \begin{bmatrix} \Sigma_{TS2SLS}^{strat} & a' \\ a & b \end{bmatrix}$$

where

$$\begin{aligned} a' = \; & E(s_{1i}w_iz_i')[E(s_{1i}z_iz_i')]^{-1}\Big(E(s_{2i}z_ix_i')\beta - E[(s_{1i} + \kappa^2 s_{2i})z_iz_i'\Pi'\beta\text{vech}(z_iz_i)'] \\ & \times \{E[(s_{1i} + \kappa^2 s_{2i})\text{vech}(z_iz_i)\text{vech}(z_iz_i)']\}^{-1}E(s_{2i}\text{vech}(z_iz_i'))\Big) \end{aligned}$$

and b is a positive number. Because a is not necessarily zero in general and b is positive, Σ_{TS2SLS}^{strat} cannot be greater than Σ_{RTSIV}^{strat} by Theorem 13 of Amemiya (1985, p.460).

References

Amemiya, Takeshi (1985), *Advanced Econometrics,* Harvard University Press: Cambridge, MA.

Andersen, Torben G. and Bent E. Sorensen (1996), "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics,* 14, 328–352.

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics,* 14, 57–67.

Angrist, Joshua D. and Alan B. Krueger (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association,* 87, 328–336.

Angrist, Joshua D. and Alan B. Krueger (1995), "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics,* 13, 225–235.

Björklund, Anders and Markus Jäntti (1997), "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review,* 87, 1009–1018.

Borjas, George J. (2004), "Food Insecurity and Public Assistance," *Journal of Public Economics,* 88, 1421–1443.

Currie, Janet and Aaron Yelowitz (2000), "Are Public Housing Projects Good for Kids?" *Journal of Public Economics,* 75, 99–124.

Dee, Thomas S. and William N. Evans (2003), "Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates," *Journal of Labor Economics,* 21, 178–209.

Hansen, Lars Peter (1982), "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica,* 50, 1029–1054.

Horn, Roger A. and Charles R. Johnson (1985), *Matrix Analysis,* Cambridge University Press: Cambridge, United Kingdom.

Murphy, Kevin M. and Robert H. Topel (1985), "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics,* 3, 88–97.

Tauchen, George (1986), "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics,* 4, 397–416.